DOCUMENT RESUME

ED 169 067 TM 007 390

AUTHOR McIntosh, Vergil M.
TITLE Estimating Examination Failure Rates and Reliability Prior to Administration.
INSTITUTION Air Univ., Gunter AFS, Ala. Extension Course Inst.
PUB DATE Oct 75
NOTE 16p.; Not available in hard copy due to marginal legibility of original document

EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS Complexity Level; Item Banks; Postsecondary Education; Statistical Analysis; Student Testing; *Test Construction; Test Items; *Test Reliability
IDENTIFIERS Item Discrimination (Tests)

ABSTRACT

Using estimates of item ease and item discrimination, procedures are provided for computing estimates of the reliability and percentage of failing scores for tests assembled from these items. Two assumptions are made: that the average item coefficient will be approximately equal to the average of the estimated coefficients and that the score distribution for the test will be approximately normal. The predicted mean test score is equal to the sum of item ease coefficients, and the predicted variance is equal to the square of the sum of item discrimination indices divided by 4.5. The fail point is always sixty percent of the number of items. Normal curve tables are used to estimate the percent of score below the fail point. A normal curve probability table is provided, as is a computer program in BASIC for using this method. The results of using this procedure with seventeen tests are presented and compared with the obtained score distribution statistics for samples of either 51 cases or 201 cases. Kuder-Richardson formula 21 reliability coefficients may also be obtained from this procedure. (CTM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED169067

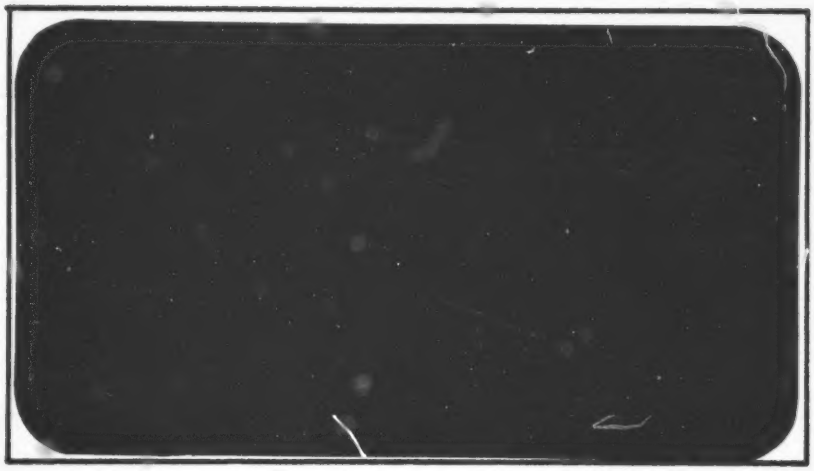
IM007 390

Extension Course Institute

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

UNITED STATES AIR FORCE
GUNTER AFS, ALABAMA



BEST COPY AVAILABLE

Professional Education is Paramount

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Virgil M. McIntosh

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

EXTENSION COURSE INSTITUTE

ESTIMATING EXAMINATION FAILURE RATES AND RELIABILITY PRIOR TO ADMINISTRATION

by

Vergil M. McIntosh

BEST COPY AVAILABLE

October 1975

Research and Evaluation Division
Extension Course Institute
Air University

FOREWORD

The USAF Extension Course Institute, with hundreds of courses and thousands of examinations, is in an excellent position to apply sophisticated techniques in its evaluation program. One such technique is described here -- a program to estimate failure rates and reliability prior to test administration.

Since the field testing and refinement of so many instruments is a luxury beyond our means, predictive measures of difficulty and reliability are necessary tools of test development and evaluation. Mr. Vergil McIntosh, of the ECI Evaluation and Research Division, has developed predictive measures that meet our needs admirably in this area.

This report on the programs he has developed has been published in the thought that other educational institutions, both military and civilian, can benefit from our findings. The comments of users would be appreciated.

HAROLD MARKOWITZ, JR., Ltcol, USAF
Chief, Evaluation and Research Division

TABLE OF CONTENTS

SECTION		PAGE
A	Introduction	1
B	Procedures Statistical Formulas Manual Computations	1
C	Computer Program for Computing Estimates ISE 2 Program for Computing Estimates Example of a Statistical Report	6
D	Conclusions Findings Summary	9

LIST OF FIGURES

FIGURE		
1	Worksheet for Computing Estimates Manually	4
2	Normal Curve Probability Table	5
3	ISE Program to Compute Examination Statistical Estimates	7
4	Printout of a Statistical Report	8

ESTIMATING EXAMINATION FAILURE RATES AND RELIABILITY PRIOR TO ADMINISTRATION

Section A - Introduction

Problem:

Because of the requirement to place examinations in use before pre-testing, the Extension Course Institute (ECI) sometimes finds that examinations are too difficult or their reliability is not high enough to be acceptable.[1] Therefore, a procedure is needed to accurately predict these test statistics before the test is activated.

To meet these needs a system has been devised and evaluated to estimate test statistics by making an estimate of the ease and discrimination index for each item. The procedure was tried, compared with actual statistical analyses, and found, in nearly all cases, to give close approximations.

The procedure was first computed manually using a worksheet and a normal curve probability table. A computer program was later developed which makes the computations and prints out a report in approximately one minute. Both the manual and the computer procedures are described in the following sections.

Section B - Procedures

Statistical Formulas:

In order to follow the rationale for the procedure it is necessary to consider the statistical formulas involved in the present statistical analysis of examinations. These formulas are:

Reliability - Kuder-Richardson Formula 21.

$$R = \frac{n\sigma^2 - M(n-M)}{\sigma^2 (n-1)}$$

Where: n = the number of items on the examination; σ = the standard deviation of scores; M = mean of examination scores.

---

1. Internal standards define an unacceptable examination as one having a failure rate in excess of 35% and/or a reliability coefficient of less than 0.75.

$$\text{Standard Deviation } (\sigma) = \sqrt{\frac{\Sigma x^2}{n}}$$

where: x = any deviation from the mean; x^2 = sum of the squared deviations; N = number of cases.

Since we do not have all of the data available to substitute in the above formulas until a sample of student solutions has been received, it is obvious that we must make some estimates.

Ebel[2] gives a formula which can be used to estimate the variance of the scores on a test. It is expressed as:

$$\sigma^2 = \frac{(\Sigma D)^2}{6}$$

where: D is the sum of the indices of discrimination for a test.

In using this formula to predict the variance of a sample of ECI tests it was found that variance can be predicted best by using a divisor of about 4.5. The reason for this difference is not known, but Ebel may have used a different formula for computing discrimination indexes.

In order to estimate the failure rate, it is necessary to compute the area under the normal probability curve falling below the fail score. This can be computed by determining the difference in standard deviations between the mean and the fail point by the formula:

$$\text{SD diff} = \frac{x}{\sigma}$$

where: x is the difference in score units between the mean and the fail score; and σ is the standard deviation of the scores. By referring to a table of the fractional parts of the area under the normal probability curve, the percent of scores falling between the mean and fail point can be determined (e.g. Table A p 458 in Garrett Statistics in Psychology and Education).

---

2. R.E. Ebel, Essentials of Educational Measurement, Englewood Cliffs, NJ: Prentice-Hall, 1972, p. 399-401.

Subtracting this value from 50 percent results in the percent of estimated failures for the examination. This, of course assumes student scores approximate a normal distribution. In using this procedure with a group of ECI courses, it was found that the predictions were generally close to the actual failure rate.

Manual Computations:

The steps in estimating the examination statistics are as follows:

STEP 1: ESTIMATE THE EASE INDEX AND DISCRIMINATION INDEX FOR EACH ITEM IN THE ITEM BANK. This step is done by the test constructor as he checks the item pool. If the items have been used on previous examinations, the item analyses statistics can provide a good basis for estimating the expected performance of each item. Estimates for individual items may not have a high degree of accuracy; however, when averages for all items are computed, the estimated and actual performance ought not differ greatly. This generalization is drawn from the known fact that a number of estimates when averaged will be very close to the true value. This step can be refined and the accuracy improved through (a) preparing guidelines for making estimates, (b) collecting and analyzing data on estimates, and (c) holding in-service training on making estimates for test constructors.

STEP 2: SELECT ITEMS FOR THE TWO PARALLEL COURSE EXAMINATIONS (CE) FORMS AND COMPUTE THE AVERAGES OF ITEM DISCRIMINATION INDEXES AND THE ITEM EASE INDEXES. A worksheet (see figure 1) has been devised to assist in making the computations.

STEP 3: COMPUTE THE VARIANCE (σ^2) AND STANDARD DEVIATION (σ). See page 2.

STEP 4: COMPUTE THE MEAN (M) OF THE RAW SCORES. M equals the number of items on the examination times the average item ease.

STEP 5: COMPUTE THE FAIL POINT. Fail point = .60 x the number of items on the examination.[4]

STEP 6: SUBTRACT THE FAIL POINT FROM THE MEAN AND DIVIDE THE DIFFERENCE BY THE STANDARD DEVIATION. This gives the difference in terms of standard deviation units.

---

4. Internal standards mandate this fail point which is based on Air Training Command resident school standards.

WORKSHEET
for estimating test failure rates and reliability

COURSE _____ CT form _____ Date _____

A. Number of items on the examination . . . . . _____

B. Sum of Discrimination Indexes . . . . . _____

C. Mean of Discrimination Indexes. [B/A] . . . . . _____

D. Sum of Item Base Indexes. . . . . _____

E. Mean of Base Indexes. . . . . [D/A] . . . . . _____

F. Mean of Raw Scores [E x A] . . . . . _____ (Mean)

G. Fail Score [A x .60] . . . . . _____

H. Estimated Variance (σ^2) [B^2/4.5] . . . . . _____

I. Estimated Standard Deviation [$\sqrt{H}$] . . . . . _____ (S.D.)

J. Difference between Mean and Fail Score [F - G] . . . _____

K. Difference "J" in terms of Standard Deviations [J/I]. _____

L. Percent of Scores between Mean and Fail Point (Refer to table of normal probability curve). . . . . _____

M. Estimated Failure Rate [.50 - L] . . . . . _____ (FR)

Estimate the test Reliability using Kuder-Richardson formula 21

$$R = \frac{n\sigma^2 - M(n-M)}{\sigma^2 (n-1)}$$

N. $n \times \sigma^2$ = [A x H] = ... _____

O. $n - M$ = [A - F] = ... _____

P. M (n-M) = [F x O] = ... _____

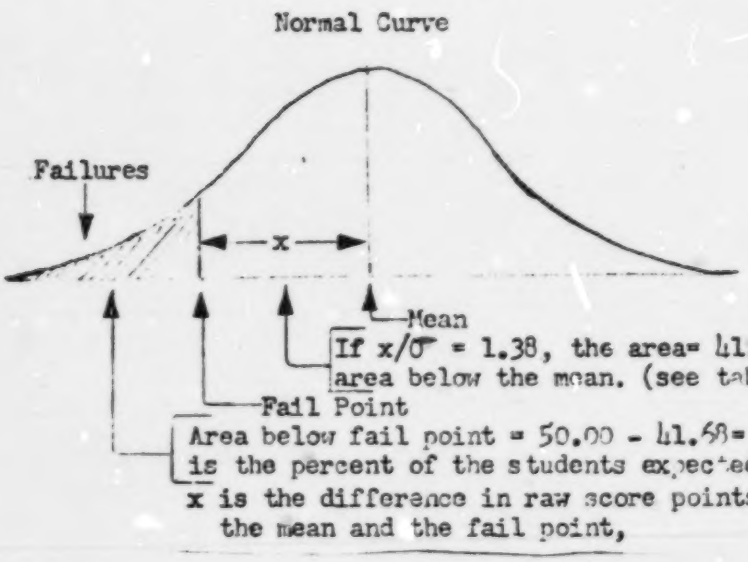
Q. The numerator = [N - P] = ... _____

R. n - 1 = [A - 1] = ... _____

S. The denominator = [H x R] = ... _____

T. Reliability = [Q/S] = ... _____ (Rel.)

Figure 1. Worksheet for Computing Estimates Manually.


Normal Curve
Failures
x
Mean
If x/σ = 1.38, the area= 41.68% of the area below the mean. (see table below)
Fail Point
Area below fail point = 50.00 - 41.68= 8.32% which is the percent of the students expected to fail exam.
x is the difference in raw score points between the mean and the fail point,

Fractional parts of the total area under the normal probability curve, corresponding to distances on the baseline between the mean and successive points laid off from the mean in units of standard deviation

Example: between the mean and a point 1.51σ ($\frac{x}{\sigma} = 1.51$) are found 43.45% of the entire area under the curve.

$\frac{x}{\sigma}$	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	0000	0040	0080	0120	0160	0199	0239	0279	0319	0359
0.1	0398	0438	0478	0517	0557	0596	0636	0675	0714	0753
0.2	0793	0832	0871	0910	0948	0987	1026	1064	1103	1141
0.3	1179	1217	1255	1293	1331	1368	1406	1443	1480	1517
0.4	1554	1591	1628	1664	1700	1736	1772	1808	1844	1879
0.5	1915	1950	1985	2019	2054	2088	2123	2157	2190	2224
0.6	2257	2291	2324	2357	2389	2422	2454	2486	2517	2549
0.7	2580	2611	2642	2673	2704	2734	2764	2794	2823	2852
0.8	2881	2910	2939	2967	2995	3023	3051	3078	3106	3133
0.9	3159	3186	3212	3238	3264	3289	3315	3340	3365	3389
1.0	3413	3438	3461	3485	3508	3531	3554	3577	3599	3621
1.1	3643	3665	3686	3708	3729	3749	3770	3790	3810	3830
1.2	3849	3869	3888	3907	3925	3944	3962	3980	3997	4015
1.3	4032	4049	4066	4082	4099	4115	4131	4147	4162	4177
1.4	4192	4207	4222	4236	4251	4265	4279	4292	4306	4319
1.5	4332	4345	4357	4370	4383	4394	4406	4418	4429	4441
1.6	4452	4463	4474	4484	4495	4505	4515	4525	4535	4545
1.7	4554	4564	4573	4582	4591	4599	4608	4616	4625	4633
1.8	4641	4649	4656	4664	4671	4678	4686	4693	4699	4706
1.9	4713	4719	4726	4732	4738	4744	4750	4756	4761	4767
2.0	4772	4778	4783	4788	4793	4798	4803	4808	4812	4817
2.1	4821	4826	4830	4834	4838	4842	4846	4850	4854	4857
2.2	4861	4864	4868	4871	4875	4878	4881	4884	4887	4890
2.3	4893	4896	4898	4901	4904	4906	4909	4911	4913	4916
2.4	4918	4920	4922	4925	4927	4929	4931	4932	4934	4936
2.5	4938	4940	4941	4943	4945	4946	4948	4949	4951	4952
2.6	4953	4955	4956	4957	4959	4960	4961	4962	4964	4964
2.7	4965	4966	4967	4968	4969	4970	4971	4972	4973	4974
2.8	4974	4975	4976	4977	4977	4978	4979	4979	4980	4981
2.9	4981	4982	4982	4983	4984	4984	4985	4985	4986	4986

Figure 2. Normal Curve Probability Table.

STEP 7: DETERMINE THE AREA UNDER THE NORMAL DISTRIBUTION CURVE BETWEEN THE MEAN AND THE FAIL POINT using the table at Figure 2.

STEP 8: SUBTRACT THE VALUE IN STEP 7 FROM 50 PERCENT. This value is the estimated failure rate. It assumes the distribution of student scores approximates a normal distribution.

STEP 9: COMPUTE THE TEST RELIABILITY by substituting the appropriate values in the reliability formula.

SECTION C - COMPUTER PROGRAM FOR COMPUTING ESTIMATES

A computer program has been written in the BASIC language to expedite the computing process. The steps in the procedure are as follows:

STEP 1: Estimate the ease and discrimination indexes for each item in the item bank.

STEP 2: Input the item ease and discrimination indexes for the selected items into a disk file via a remote terminal. Do not use decimal points in inputting the data.

STEP 3: Use the ISE 2 computer program (see Figure 3) to compute the estimates and print out a report. In running the ISE 2 program, the file name for the data file should be entered in line 060. Line 070 should be checked (listed) to assure the read statement corresponds to data listed in the file. The value "y" will read the ease index, and "z" the discrimination index.

STEP 4: A report will be printed out on the remote terminal. A sample report is shown in figure 4.

```
10 REM***THIS PROGRAM ESTIMATES MEANS,FAILURE RATES,
20 REM    AND RELIABILITY***
30 REM ***DATA IS ENTERED FROM A FILE***
40 PRINT "ENTER COURSE AND FORM NUMBER"
50 INPUT C1,C2
60 FILES 631505B
70 FILES NRMCRV1
80 READ #1,X,Y,Z,W
90 N=N+1
100 E=E+Y
110 D=D+Z
120 IF MORE #1 THEN 80
130 REM***COMPUTE AVG EASE***
140 h=e/(n*100)
150 REM*** COMPUTE AVG ITEM DISC***
160 g=(d/100)/(n)
170 REM***COMPUTE MEAN OF RAW SCORES***
180 r=n*h
190 q=n*.60
200 REM***COMPUTE VARIANCE***
210 a=4.5
220 v=(d/100)^2
230 v=((d/100)^2)/(a)
240 s=v^.5
250 rem compute diff mean and fail pt in sd
260 o=(r-q)/s
270 print "diff mean and fp in sd=",o
280 O=(O*10)+.5 \O=INT(O)
290 READ #2,E,F
300 IF E<>0 GOTO290
310 I=.50 -F
320 k=(n*v-r*(n-r))/(v*(n-1))
330 PRINT\PRINT\PRINT
340 print tab(14);"COURSE EXAMINATION STATISTICAL ESTIMATES"
350 PRINT TAB(16);"COURSE";C1     "FORM";C2;SPC(10)"DATE"; SPC(2);DAT$
360 print using 370,    n
370: NR ITEMS=                  ###
380 print using 390,            h
390: AVG EASE=                 .###
400 print using 410,   g
410: AVG ITEM DISC=            .###
420 print using 430,   r
430: MEAN=                      ##.#
440 print using 450,   s
450: STANDARD DEVIATION=       ##.##
460 print using 470,q
470: PASS/FAIL POINT=           ##.#
480 print using 490,   t
490: EST FAILURE RATE=           .##
500 print using 510,   k
510: RELIABILITY=               #.###
520 print
530 END
```

FIGURE 3. ISE PROGRAM TO COMPUTE EXAMINATION STATISTICAL ESTIMATE

* 7

COURSE EXAMINATION STATISTICAL ESTIMATES

COURSE 63150 FORM 25 DATE 10/06/75

NR ITEMS=	72
AVG EASE=	.710
AVG ITEM DISC=	.281
MEAN=	51.1
STANDARD DEVIATION=	9.53
PASS/FAIL POINT=	43.2
EST FAILURE RATE=	.20
RELIABILITY=	.849

Figure 4. Printout of a Statistical Report.

Section D - Conclusions

Findings:

Comparisons were made between the estimates for several courses and item analyses based on samples of student test papers. The results showed generally close agreement. Differences were approximately of the same magnitude as differences found between two different analyses. Figure 5 is a table comparing estimates with student samples of 51 and 201. Zeros on the table indicate that data are not available. ("CRSE" and "FM" indicate ECI course and examination form number.

COURSE EXAMINATION STATISTICS

CRSE	FM	AVG EASE			AVG DISC			STD DEV			FAIL RATE			RELIABILITY		
		EST	51	201	EST	51	201	EST	51	201	EST	51	201	EST	51	201
32371	24	.74	.71	.00	.29	.21	.00	8.5	5.9	.0	.15	.14	.00	.86	.64	.00
54211	27	.73	.71	.73	.37	.39	.36	15.1	14.6	.0	.13	.31	.24	.93	.93	.92
54211	28	.76	.73	.74	.34	.34	.33	14.1	13.0	.0	.17	.23	.20	.93	.91	.91
54550	35	.73	.68	.67	.29	.32	.32	12.7	13.8	.0	.25	.29	.32	.89	.91	.90
54550	36	.73	.69	.69	.29	.26	.29	13.0	10.8	.0	.24	.22	.23	.90	.85	.88
54750	25	.78	.66	.66	.29	.24	.27	13.3	10.4	.0	.09	.33	.31	.91	.81	.85
54750	26	.77	.59	.64	.31	.29	.31	14.3	12.5	.0	.12	.57	.36	.93	.86	.88
55254	27	.72	.71	.00	.38	.33	.00	16.9	.0	.0	.25	.24	.00	.94	.91	.00
55254	28	.72	.77	.00	.33	.25	.00	15.1	.0	.0	.22	.08	.00	.93	.86	.00
62150	25	.75	.67	.00	.29	.24	.00	9.7	8.3	.0	.13	.22	.00	.84	.78	.00
63130	24	.75	.72	.75	.33	.31	.27	13.8	12.4	11.6	.16	.18	.12	.92	.89	.89
63130	25	.73	.73	.74	.31	.32	.27	13.2	12.4	11.3	.18	.18	.11	.91	.90	.88
63150	25	.72	.72	.74	.34	.26	.27	12.0	9.1	8.9	.22	.18	.16	.90	.83	.84
63150	26	.73	.70	.72	.32	.25	.27	11.3	9.2	9.4	.19	.14	.16	.89	.82	.84
64552	24	.74	.76	.00	.34	.36	.00	9.6	.0	.0	.19	.18	.00	.88	.92	.00
64552	23	.74	.72	.00	.34	.30	.00	8.8	.0	.0	.19	.12	.00	.89	.86	.00
90650	26	.72	.70	.73	.28	.19	.26	13.4	8.8	7.2	.18	.12	.12	.90	.74	.86

Figure 5. Comparison of Estimated Statistics with Analysis of Student Samples.

Although estimates are generally close to the actual analyses, it is likely that some refinements can be made to the procedure, and guidelines can be prepared to assist test constructors in making estimates, and thus improve these estimates.

Significant advantages to be realized from using the estimating procedure are that it will (1) help assure that different forms of the CEs are equivalent, (2) reduce the number of CEs with excessive failure rates or low reliability and (3) require test constructors to carefully evaluate an item's function in a test. This will result in distinct improvement in test quality.

Summary:

A system has been developed to estimate examination statistics before the examination has been administered. The system requires the test constructor to make an estimate of the ease index and discrimination index for each item. These indexes are then used to compute test estimates using the worksheet or the computer program. Based on samples of actual student data the system has been found to provide relatively close estimates of test performance.

REFERENCES

Ebel, Robert L., Essentials of Educational Measurement, Englewood Cliffs, NJ: Prentice-Hall, 1972.

Garrett, Henry E., Statistics in Psychology and Education New York: David McKay Co. Inc., 1966.

Extension Course Institute, Hq Operating Instruction 11-12, Volume Review Exercises and Course Examinations, 20 October 1972.